FAST LAPPED IMAGE TRANSFORMS USING LIFTING STEPS

FIELD OF THE INVENTION

The current invention relates to the processing of images such as photographs, drawings, and other two dimensional displays. It further relates to the processing of such images which are captured in digital format or after they have been converted to or expressed in digital format. This invention further relates to use of novel coding methods to increase the speed and compression ratio for digital image storage and transmission while avoiding introduction of undesirable artifacts into the reconstructed images.

BACKGROUND OF THE INVENTION

In general, image processing is the analysis and manipulation of two-dimensional representations, which can comprise photographs, drawings, paintings, blueprints, x-rays of medical patients, or indeed abstract art or artistic patterns. These images are all two-dimensional arrays of information. Until fairly recently, images have comprised almost exclusively analog displays of analog information, for example, conventional photographs and motion pictures. Even the signals encoding television pictures, notwithstanding that the vertical scan comprises a finite number of lines, are fundamentally analog in nature.

Beginning in the early 1960's, images began to be captured or converted and stored as two-dimensional digital data, and digital image processing followed. At first, images were recorded or transmitted in analog form and then converted to digital representation for manipulation on a computer. Currently digital capture and transmission are on their way to dominance, in part because of the advent of charge coupled device (CCD) image recording arrays and in part because of the availability of inexpensive high speed computers to store and manipulate images.

An important task of image processing is the correction or enhancement of a particular image. For example, digital enhancement of images of celestial objects taken by space probes has provided substantial scientific information. However, the current invention relates primarily to compression for transmission or storage of digital images and not to enhancement.

One of the problems with digital images is that a complete single image frame can require up to several megabytes of storage space or transmission bandwidth. That is, one of today's 3½ inch floppy discs can hold at best a little more than one gray-scale frame and sometimes substantially less than one whole frame. A full-page color picture, for example, uncompressed, can occupy 30 megabytes of storage space. Storing or transmitting the vast amounts of data which would be required for real-time uncompressed high resolution digital video is technologically daunting and virtually impossible for many important communication channels, such as the telephone line. The transmission of digital images from space probes can take many hours or even days if insufficiently compressed images are involved. Accordingly, there has been a decades long effort to develop methods of extracting from images the information essential to an aesthetically pleasing or scientifically useful picture without degrading the image quality too much and especially without introducing unsightly or confusing artifacts into the image.

The basic approach has usually involved some form of coding of picture intensities coupled with quantization. One approach is block coding; another approach, mathematically equivalent with proper phasing, is multiphase filter banks. Frequency based multi-band transforms have long found application in image coding. For instance, the JPEG image compression standard, W. B. Pennebaker and J. L. Mitchell, "JPEG: Still Image Compression Standard," Van Nostrand Reinhold, 1993, employs the 8x8 discrete cosine transform (DCT) at its transformation stage. At high bit rates, JPEG offers almost lossless reconstructed image quality. However, when more compression is needed, annoying blocking artifacts appear since the DCT bases are short and do not overlap, creating discontinuities at block boundaries.

The wavelet transform, on the other hand, with long, varying-length, and overlapping bases, has elegantly solved the blocking problem. However, the transform's computational complexity can be significantly higher than that of the DCT. This complexity gap is partly in terms of the number of arithmetical operations involved, but more importantly, in terms of the memory buffer space required. In particular, some implementations of the wavelet transform require many more operations per output coefficient as well as a large buffer.

An interesting alternative to wavelets is the lapped transform, e.g., H. S. Malvar, Signal Processing with Lapped Transforms, Artech House, 1992, where pixels from adjacent blocks are utilized in the calculation of transform coefficients for the working block. The lapped transforms outperform the DCT on two counts: (i) from the analysis viewpoint, they take into account inter-block correlation and hence provide better energy compaction; (ii) from the synthesis viewpoint, their overlapping basis functions decay asymptotically to zero at the ends, reducing blocking discontinuities dramatically.

Nevertheless, lapped transforms have not yet been able to supplant the unadorned DCT in international standard coding routines. The principal reason is that the modest improvement in coding performance available up to now has not been sufficient to justify the significant increase in computational complexity. In the prior art, therefore, lapped transforms remained too computationally complex for the benefits they provided. In particular, the previous lapped transformed somewhat reduced but did not eliminate the annoying blocking artifacts.

It is therefore an object of the current invention to provide a new transform which is simple and fast enough to replace the bare DCT in international standards, in particular in JPEG and MPEG-like coding standards. It is another object of this invention to provide an image transform which has overlapping basis functions so as to avoid blocking artifacts. It is a further object of this invention to provide a lapped transform which is approximately as fast as, but more efficient for compression than, the bare DCT. It is yet another object of this invention to provide dramatically improved speed and efficiency using a lapped transform with lifting steps in a butterfly structure with dyadic-rational coefficients. It is yet a further object of this invention to provide a transform structure such that for a negligible complexity surplus over the bare DCT a dramatic coding performance gain can be obtained both from a subjective and objective point of view while blocking artifacts are completely eliminated.

SUMMARY OF THE INVENTION

In the current invention, we use a family of lapped biorthogonal transforms implementing a small number of dyadic-rational lifting steps. The resulting transform, called the LiftLT, not only has high computation speed but is well-suited to implementation via VLSI.

Moreover, it also consistently outperforms state-of-the-art wavelet based coding systems in coding performance when the same quantizer and entropy coder are used. The LiftLT is a lapped biorthogonal transform using lifting steps in a modular lattice structure, the result of which is a fast, efficient, and robust encoding system. With only 1 more multiplication (which can also be implemented with shift-and-add operations), 22 more additions, and 4 more delay elements compared to the bare DCT, the LiftLT offers a fast, low-cost approach capable of straightforward VLSI implementation while providing reconstructed images which are high in quality, both objectively and subjectively. Despite its simplicity, the LiftLT provides a significant improvement in reconstructed image quality over the traditional DCT in that blocking is completely eliminated while at medium and high compression ratios ringing artifacts are reasonably contained. The performance of the LiftLT surpasses even that of the well-known 9/7-tap biorthogonal wavelet transform with irrational coefficients. The LiftLT's block-based structure also provides several other advantages: supporting parallel processing mode, facilitating region-of-interest coding and decoding, and processing large images under severe memory constraints.

Most generally, the current invention is an apparatus for block coding of windows of digitally represented images comprising a chain of lattices of lapped transforms with dyadic rational lifting steps. More particularly, this invention is a system of electronic devices which codes, stores or transmits, and decodes M×M sized blocks of digitally represented images, where M is an even number. The main block transform structure comprises a transform having M channels numbered 0 through M−1, half of said channel numbers being odd and half being even; a normalizer with a dyadic rational normalization factor in each of said M channels; two lifting steps with a first set of identical dyadic rational coefficients connecting each pair of adjacent numbered channels in a butterfly configuration, M/2 delay lines in the odd numbered channels; two inverse lifting steps with the first set of dyadic rational coefficients connecting each pair of adjacent numbered channels in a butterfly configuration; and two lifting steps with a second set of identical dyadic rational coefficients connecting each pair of adjacent odd numbered channels; means for transmission or storage of the transform output coefficients; and an inverse transform comprising M channels numbered 0 through M−1, half of said channel numbers being odd and half being even; two inverse lifting steps with dyadic rational coefficients connecting each pair of adjacent odd numbered channels; two lifting steps with dyadic rational coefficients connecting each pair of adjacent numbered channels in a butterfly configuration; M/2 delay lines in the even numbered channels; two inverse lifting steps with dyadic rational coefficients connecting each pair of adjacent numbered channels in a butterfly configuration; a denormalizer with a dyadic rational inverse normalization factor in each of said M channels; and a base inverse transform having M channels numbered 0 through M−1.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a polyphase representation of a linear phase perfect reconstruction filter bank.

FIG. 2 shows the most general lattice structure for linear phase lapped transforms with filter length L=KM.

FIG. 3 shows the parameterization of an invertible matrix via the singular value decomposition.

FIG. 4 portrays the basic butterfly lifting configuration.

FIG. 5 depicts the analysis LiftLT lattice drawn for M=8.

FIG. 6 depicts the synthesis LiftLT lattice drawn for M=8.

FIG. 7 depicts a VLSI implementation of the analysis filter bank operations.

FIG. 8 shows frequency and time responses of the 8×16 LiftLT: Left: analysis bank. Right: synthesis bank.

FIG. 9 portrays reconstructed "Barbara" images at 1:32 compression ratio.

DESCRIPTION OF THE PREFERRED EMBODIMENT

Typically, a block transform for image processing is applied to a block (or window) of, for example, 8×8 group of pixels and the process is iterated over the entire image. A biorthogonal transform in a block coder uses as a decomposition basis a complete set of basis vectors, similar to an orthogonal basis. However, the basis vectors are more general in that they may not be orthogonal to all other basis vectors. The restriction is that there is a "dual" basis to the original biorthogonal basis such that every vector in the original basis has a "dual" vector in the dual basis to which it is orthogonal. The basic idea of combining the concepts of biorthogonality and lapped transforms has already appeared in the prior art. The most general lattice for M-channel linear phase lapped biorthogonal transforms is presented in T. D. Tran, R. de Queiroz, and T. Q. Nguyen, "The generalized lapped biorthogonal transform," ICASSP, pp. 1441–1444, Seattle, May 1998, and in T. D. Tran, R. L. de Queiroz, and T. Q. Nguyen, "Linear phase perfect reconstruction filter bank: lattice structure, design, and application in image coding" (submitted to EEE Trans. on Signal Processing, April 1998). A signal processing flow diagram of this well-known generalized filter bank is shown in FIG. 2.

In the current invention, which we call the Fast LiftLT, we apply lapped transforms based on using fast lifting steps in an M-channel uniform linear-phase perfect reconstruction filter bank, according to the generic polyphase representation of FIG. 1. In the lapped biorthogonal approach, the polyphase matrix E(z) can be factorized as

$$E(z)=G_{K-1}(z)G_{K-2}(z)\ldots G(z)E_0(z), \tag{1}$$

where

$$G_i(z)=\frac{1}{2}\begin{bmatrix} U_i & 0 \\ 0 & V_i \end{bmatrix}\begin{bmatrix} I & I \\ I & -I \end{bmatrix}\begin{bmatrix} I & 0 \\ 0 & z^{-1}I \end{bmatrix}\begin{bmatrix} I & I \\ I & -I \end{bmatrix} \tag{2}$$

$$\equiv\frac{1}{2}\Phi_i\times W\times\Lambda\times W, \text{ and}$$

$$E_0(z)=\frac{1}{\sqrt{2}}\begin{bmatrix} U_0 & U_0J_{M/2} \\ V_0J_{M/2} & -V_0 \end{bmatrix}. \tag{3}$$

In these equations, I is the identity matrix, and J is the matrix with 1's on the anti-diagonal.

The transform decomposition expressed by equations (1) through (3) is readily represented, as shown in FIG. 2, as a complete lattice replacing the "analysis" filter bank E(z) of FIG. 1. This decomposition results in a lattice of filters having length L=KM. (K is often called the overlapping factor.) Each cascading structure $G_1(z)$ increases the filter length by M. All U_i and V_i, i=0,1, . . . , K−1, are arbitrary M/2×M/2 invertible matrices. According to a theorem well known in the art, invertible matrices can be completely represented by their singular value decomposition (SVD), given by

$$U_i=U_{i0}\Gamma_iU_{i1},\ V_i=V_{i0}\Delta_iV_{i1}$$

where U_{i0}, U_{i1}, V_{i0}, V_{i1} are diagonalizing orthogonal matrices and Γ_i, Δ_i are diagonal matrices with positive elements.

It is well known that any M/2×M/2 orthogonal matrix can be factorized into M(M−2)/8 plane rotations θ_i and that the diagonal matrices represent simply scaling factors α_i. Accordingly, the most general LT lattice consists of KM(M−2)/2 two dimensional rotations and 2M diagonal scaling factors α_i. Any invertible matrix can be expressed as a sequence of pairwise plane rotations θ and scaling factors α_i as shown in FIG. 3.

It is also well known that a plane rotation can be performed by 3 "shears":

$$\begin{bmatrix}\cos\theta_i & -\sin\theta_i\\ \sin\theta_i & \cos\theta_i\end{bmatrix}=\begin{bmatrix}1 & \frac{\cos\theta_i-1}{\sin\theta_i}\\ 0 & 1\end{bmatrix}\begin{bmatrix}1 & 0\\ \sin\theta_i & 1\end{bmatrix}\begin{bmatrix}1 & \frac{\cos\theta_i-1}{\sin\theta_i}\\ 0 & 1\end{bmatrix}.$$

This can be easily verified by computation

Each of the factors above is capable of a "lifting" step in signal processing terminology. The product of two which effects a linear transform of pairs of coefficients:

$$\begin{bmatrix}a\\ b\end{bmatrix}\to\begin{bmatrix}1+km & k\\ m & 1\end{bmatrix}\times\begin{bmatrix}a\\ b\end{bmatrix}.$$

The signal processing flow diagram of this operation is shown in FIG. 4. The crossing arrangement of these flow paths is also referred to as a butterfly configuration. Each of the above "shears" can be written as a lifting step.

Combining the foregoing, the shears referred to can be expressed as computationally equivalent "lifting steps" in signal processing. In other words, we can replace each "rotation" by 3 closely-related lifting steps with butterfly structure. It is possible therefore to implement the complete LT lattice shown in FIG. 2 by 3KM(−2)/2 lifting steps and 2M scaling multipliers.

In the simplest but currently preferred embodiment, to minimize the complexity of the transform we choose a small overlapping factor K=2 and set the initial stage E_0 to be the DCT itself Many other coding transforms can serve for the base stage instead of the DCT, and it should be recognized that many other embodiments are possible and can be implemented by one skilled in the art of signal processing.

Following the observation in H. S. Malvar, "Lapped biorthogonal transforms for transform coding with reduced blocking and ringing artifacts," ICASSP97, Munich, April 1997, we apply a scaling factor to the first DCT's antisymmetric basis to generate synthesis LT basis functions whose end values decay smoothly to exact zero—a crucial advantage in blocking artifacts elimination. However, instead of scaling the analysis by $\sqrt{2}$ and the synthesis by $1/\sqrt{2}$, we opt for 25/16 and its inverse 16/25 since they allow the implementation of both analysis and synthesis banks in integer arithmetic. Another value that works almost as well as 25/16 is 5/4. To summarize, the following choices are made in the first stage: the combination of U_{00} and V_{00} with the previous butterfly form the DCT;

$$\Delta_0=diag\left[\frac{25}{16},1,\ldots,1\right],$$

and $\Gamma_0=U_{00}=V_{00}=I_{M/2}$. See FIG. 2.

After 2 series of ±1 butterflies W and the delay chain $\Lambda(z)$, the LT symmetric basis functions already have good attenuation, especially at DC ($\omega=0$). Hence, we can comfortably set $U_1=I_{M/2}$.

As noted, V_1 is factorizable into a series of lifting steps and diagonal scalings. However, there are several problems: (i) the large number of lifting steps is costly in both speed and physical real-estate in VLSI implementation; (ii) the lifting steps are related; (iii) and it is not immediately obvious what choices of rotation angles will result in dyadic rational lifting multipliers. In the current invention, we approximate V_1 by (M/2)−1 combinations of block-diagonal predict-and-update lifting steps, i.e.,

$$\begin{bmatrix}1 & u_i\\ 0 & 1\end{bmatrix}\times\begin{bmatrix}1 & 0\\ -p_i & 1\end{bmatrix}.$$

Here, the free parameters u_i and p_i can be chosen arbitrarily and independently without affecting perfect reconstruction. The inverses are trivially obtained by switching the order and the sign of the lifting steps. Unlike popular lifting implementations of various wavelets, all of our lifting steps are of zero-order, namely operating in the same time epoch. In other words, we simply use a series of 2×2 upper or lower diagonal matrices to parameterize the invertible matrix V_1.

Most importantly, fast-computable VLSI-friendly transforms are readily available when u_i and p_i are restricted to dyadic rational values, that is, rational fractions having (preferably small) powers of 2 denominators. With such coefficients, transform operations can for the most part be reduced to a small number of shifts and adds. In particular, setting all of the approximating lifting step coefficients to −1/2 yields a very fast and elegant lapped transform. With this choice, each lifting step can be implemented using only one simple bit shift and one addition.

The resulting LiftLT lattice structures are presented in FIGS. 5 and 6. The analysis filter shown in FIG. 5 comprises a DCT block 1, 25/16 normalization 2, a delay line 3 on four of the eight channels, a butterfly structured set of lifting steps 5, and a set of four fast dyadic lifting steps 6. The frequency and impulse responses of the 8×16 LiftLT's basis functions are depicted in FIG. 8.

The inverse or synthesis lattice is shown in FIG. 6. This system comprises a set of four fast dyadic lifting steps 11, a butterfly-structured set of lifting steps 12, a delay line 13 on four of the eight channels, 16/25 inverse normalization 14, and an inverse DCT block 15. FIG. 7 also shows the frequency and impulse responses of the synthesis lattice.

The LiftLT is sufficiently fast for many applications, especially in hardware, since most of the incrementally added computation comes from the 2 butterflies and the 6 shift-and-add lifting steps. It is faster than the type-I fast LOT described in H. S. Malvar, *Signal Processing with Lapped Transforms,* Artech House, 1992. Besides its low complexity, the LiftLT possesses many characteristics of a high-performance transform in image compression: (i) it has high energy compaction due to a high coding gain and a low attenuation near DC where most of the image energy is concentrated; (ii) its synthesis basis functions also decay smoothly to zero, resulting in blocking-free reconstructed images.

Comparisons of complexity and performance between the LiftLT and other popular transforms are tabulated in Table 1 and Table 2. The LiftLT's performance is already very close to that of the optimal generalized lapped biorthogonal transform, while its complexity is the lowest amongst the transforms except for the DCT.

To assess the new method in image coding, we compared images coded and decoded with four different transforms:

DCT: 8-channel, 8-tap filters

Type-I Fast LOT: 8-channel, 16-tap filters

LiftLT: 8-channel, 16-tap filters

Wavelet: 9/7-tap biorthogonal.